

ユーザーとのインタラクションによる心象風景の可視化モデルの提案

千葉 幹也

指導教員：小林 邦和

1 はじめに

近年、深層学習の分野は急激な技術革新が進んでいる。深層学習の発展として、脳波をfMRIを用いて測定し、それをDeep Neural Networkに学習させ、脳波から心象風景の可視化を行うという研究もおこなわれている[1]。この研究の利点は、直接脳波から情報を得るので、得られる情報量が多く、物体の位置情報や形状情報などを含めた画像が生成可能である点である。欠点として、脳波を測定する必要があるため、脳波測定用の機器を装着しなければならない点、脳波特徴と画像特徴を同一空間上に描写し、特徴が近い画像が生成されるように繰り返し画像の最適化を行うため、実際に画像として表示されるまでに時間がかかる点があげられる。

卒業研究[2]では、脳波ではなく説明文章を入力データとすることで、先行研究であるfMRIを用いた心象風景の可視化の欠点である、脳波を扱わなければならない点の解決を目指した。提案したモデルは、説明文章を要約するため、要約モデルとしてPointer-Generator Networks[3]と短いキャプションから画像生成を行うAttentional Generative Adversarial Networks[4]を組み合わせたものである。提案したモデルで計算機シミュレーションを行った結果、生成される画像の精度は犬や別種の鳥よりも元画像に近い精度という結果となったが、中には人の目から見ると明らかに鳥ではない画像が生成されてしまう問題があった。

また、文章要約時に情報が欠落してしまう問題や、ユーザーから得られる情報が最初の文章しかないため、本当にユーザーが求めている画像なのかどうか判断する材料が少ないという問題もあった。

そこで本研究では、卒業研究で提案したモデル(以下、従来モデルと表記)の問題点の解決を行い、従来モデルの改善を行うことを目的とし、最終目標として実用化が容易であり、高精度な心象風景の可視化モデルの構成を目指す。

2 提案モデル

本研究ではユーザーとのインタラクションを行うことで、従来モデルと比較し、よりユーザーの求める画像が生成されるようなモデルの構築を行う。インタラクションには様々な方法が存在するが、本研究では生成された画像からユーザーに選択してもらった部分に関して後述の2種類のモデルを提案する。

2.1 提案モデル1

提案モデル1では、AttnGANでは行うことができない後から情報を追加できる機能を追加するため、文章特徴ベクトルと単語特徴ベクトルを用いて、次の画像生成時に一つ前の特徴ベクトルを用いる機構を追加した。また、初回の生成時に、複数のシード値から生成を行い、ユーザーが選択できることにすることで、明らかに鳥ではない画像が生成されることを回避できるようにした。提案モデル1の構成を図1に示す。

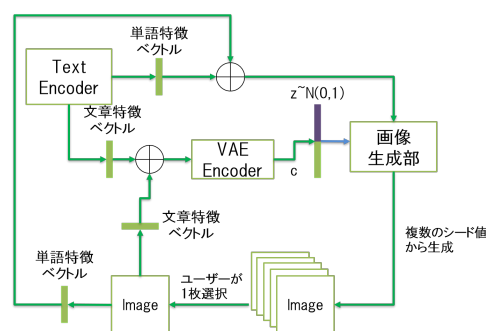


図1 提案モデル1の構成

2.2 提案モデル2

提案モデル2では、AttnGANの一部に使用されているVariational Autoencoder[5]の機能を用いて、一度に4つの文章を入力し、それを元にモーフィング画像の生成を行い、各文章の重要度をユーザーの選択により情報として入力する機構を追加した。提案モデル2の構成を図2に示し、実際に生成されるモーフィング画像の例を図3に示す。

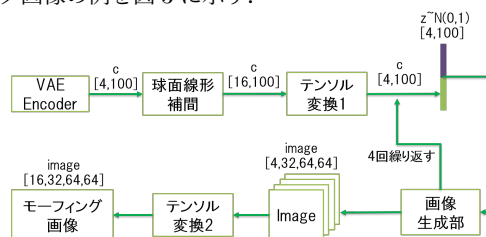


図2 提案モデル2の構成

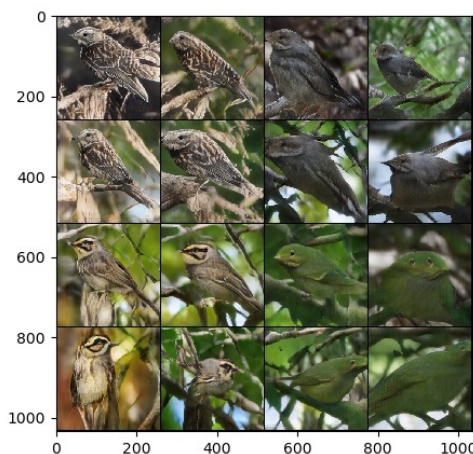


図3 モーフィング画像の例

3 計算機シミュレーション

2つの提案モデルと従来モデルの精度比較を行うため、条件を設定し、定量的な数値による評価としてBag of Visual Wordsという評価指標と、主観的評価として実際に人間に評価してもらうという2つの観点でシミュレーションを行った。

3.1 条件設定

本研究では、従来モデルとの精度比較を行うため、従来モデルのデータは条件設定をそのまま流用する。

従来モデルで用いた条件設定は図4の通りである。

- 対話文章にしない。
- 時間軸に関する記述をしない。
- 色、姿かたちについての情報を必ず説明する。
- 感情など画像生成上不必要な情報を可能であれば含める。
- 正解ラベルとして要約文章も記述する。

図4 元となった画像に対する原文作成条件

また、本研究で提案した2つモデルで用いる条件設定は図5の通りである。

- 各文章に対して生成を実行するのは1回のみ
- 文章は従来モデルで使ったものを流用(比較のため)
- 提案モデル1は文章が段階的に増えていくため、その途中結果についても評価する

図5 提案モデルの条件

1つ目の条件は、複数回実行しなおして、良い画像を作為的に選べるようにしないようにするため。3つ目の条件は提案モデル1に関しては任意の文章を読み込んだ段階で、情報を追加するかどうかユーザーが判断できるという提案モデルであるためである。

本シミュレーションを行うため10種類の鳥の説明文章に対して、各モデル10枚ずつ生成し、従来モデルは100枚、提案モデル1は100枚、提案モデル2は100枚の計300枚を用意した。

3.2 定量的評価

定量的評価では、生成画像そのまま使用した場合と、背景除去を行った場合の2種類で評価を行った。また、Bag of Visual Wordsは10回行い、その平均値を利用した。

3.2.1 背景除去をしない場合

背景除去をしない場合の類似度比較の結果は表1の通りである。表内のTop1はそのモデルが、他の2つのモデルと比較し一番精度が高かった回数を指す。

表1 背景除去をしない場合の類似度比較

	Top1	従来モデルより良い数値回数
提案モデル1	8/10	9/10
提案モデル2	2/10	7/10
従来モデル	0/10	

3.2.2 背景除去をする場合

背景除去をする場合の類似度比較の結果は表1の通りである。

表2 背景除去をする場合の類似度比較

	Top1	従来モデルより良い数値回数
提案モデル1	6/10	8/10
提案モデル2	4/10	8/10
従来モデル	0/10	

3.3 主観的評価

主観的評価では、被験者に2種類の鳥について3つのモデルの生成画像を見せ、1は最も似ていない、5は最も似ているとして5段階評価で似ているかどうかの判定を行った。5段階評価における結果を表3に示す。

表3 人間による評価

種類	モデル	A	B	C	D	E	平均
カラス	従来モデル	3	2	4	1	3	2.6
	提案モデル1	2	3	3	3	1	2.4
	提案モデル2	4	4	4	4	4	4
夜鷹	従来モデル	1	2	3	1	3	2
	提案モデル1	4	3	3	3	4	3.4
	提案モデル2	3	2	3	4	4	3.2

3.4 考察

定量的評価では提案モデル1が一番精度が高いという結果となった。要因として、文章を後から追加することが大きく影響していると考えられる。実際に1枚目生成時点で各モデルより高い精度を出した回数は表4の通りとなっている。

表4 提案モデル1の1枚目の精度比較

	従来モデル	提案モデル2
提案モデル1	6/10	3/10
提案モデル1+ 背景除去	6/10	2/10

また、主観的評価では、カラスでは提案モデル2が夜鷹では提案モデル1が一番精度が高いという結果となった。類似に関する感覚は個人差があり、形状や位置、色や特徴など人によって重要視する点も異なるため、このように評価が分かれたのだと考えられる。この結果に関しては被験者実験のデータ数が足りていないため、より多くのデータを取得していく必要がある。

4 まとめと今後の展望

本研究では、AttnGANにユーザーとのインタラクション機能を追加することで、従来モデルよりも精度向上につながるという結果が得られた。今後の展望として、追加する文章が生成画像に与える影響力の調査と、その情報を用いた新たなインタラクション機能の構築、チャットボット形式でリアルタイムに返答できる機能の実装などが挙げられる。

参考文献

- [1] Shen, G., Horikawa, T., Majima, K., Kamitani, Y. "Deep image reconstruction from human brain activity", Plos Comput. Biol. 15, e1006633, 2019.
- [2] 千葉 幹也, 小林 邦和, 鈴木 拓央: 「深層学習による会話文章の要約を用いた心象風景の可視化」 2019.
- [3] Abigail See, Peter J. Liu, Christopher D. Manning. "Get To The Point: Summarization with Pointer-Generator Networks", Association for Computational Linguistics (ACL)., 2017.
- [4] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He., "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks", CVPR, 2018.
- [5] Diederik P. Kingma, Max Welling. "Auto-Encoding Variational Bayes" ICLR. 2014.